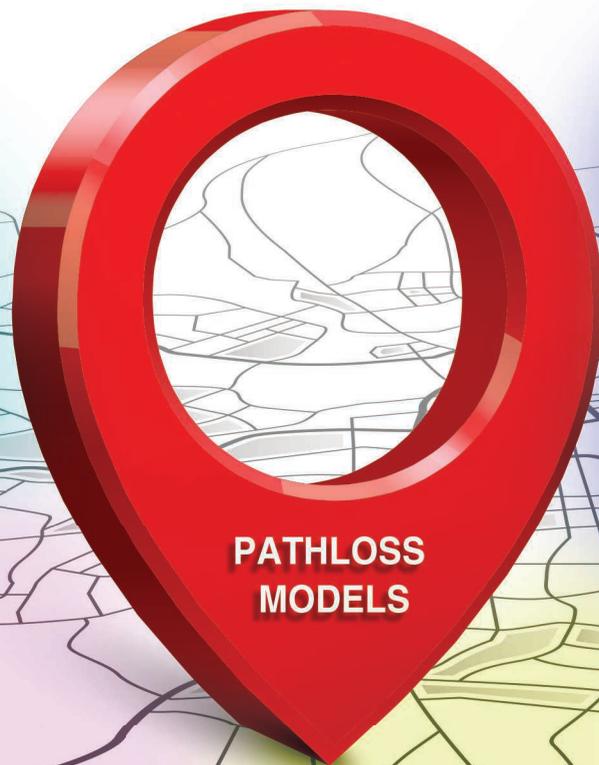


*Finding a way
to determine similarity
between areas.*

On the Cross-Application of Calibrated Pathloss Models Using Area Features



Pathloss-model calibration, or tuning, is the practice of refining the nominal parameters of a model according to measurement samples collected in a specific area. It is widely used by mobile providers because it can reduce error up to tens of decibels depending on the model category. It comes, however, at the expense of both time and monetary resources. Given the prohibitive amount of resources required to calibrate all deployment areas, a model calibrated in one area can be applied

to another area that has no measurement data. We refer to this practice as model *cross-application*. How well the model predicts will depend on the similarity between the two areas.

In this article, we propose a methodology for cross-application in which we identify the most effective features to determine area similarity. To do so, we analyzed over 3 million measurement samples from five metropolitan regions throughout the United States—consisting of urban, suburban, and rural environments—while considering a broad range of model categories, from purely empirical to highly deterministic. We also validated the



©ISTOCKPHOTO.COM/TETIANA LAZUNOVA

performance of the models per environment in terms of both absolute prediction error and error reduction effected by calibration.

INTRODUCTION

The objective of pathloss-model calibration is to have the model predictions fit the samples as well as possible, such that the model delivers more accurate results throughout the area as a whole. Depending on the model category, calibration can impact results significantly, reducing prediction error anywhere from 3 dB to tens of decibels [1]–[10]. To achieve such performance in cross-application, mobile providers will require no fewer than 300 representative models across the United States [11]. The cost of obtaining the set is on the order of several millions of dollars [11].

Prediction accuracy will depend mainly on three factors:

- **Environment:** There is greater propagation uncertainty in complex urban environments with tall man-made structures and other such clutter, in contrast to rural environments with mostly open areas. Thus, accuracy will generally be better in the latter than in the former.
- **Geodata:** Geographic data, abbreviated as *geodata*, are prior information describing the features of an environment, such as terrain and clutter. Complex environments will require a more detailed description to yield

the same level of accuracy as rural environments [12]. For example, knowing the building morphologies in an urban environment is critical to predicting shadowed zones or regions subject to waveguiding, whereas in rural environments, the clutter profile is essentially flat.

- **Model category:** Purely empirical models are derived from prior measurements in a specific area and, since they do not exploit geodata, have no perception of the actual prediction area. They typically do well in rural and some suburban environments where the landscape is similar between different areas. By contrast, highly deterministic models exploit all available geodata and so are more suitable for urban environments. However, the high-resolution geodata required to deliver accurate results can cost up to a half-million dollars [11], and prediction will take much longer than with purely empirical models.

Just as prediction accuracy depends on many factors, so does the benefit of calibration. Because empirical models are less deterministic—they exploit fewer geodata—they will rely more on measurement data and so will benefit the most from tuning. Conversely, while deterministic models do well when geodata are available, they will also see some benefit from tuning, as geodata cannot entirely capture all of the features of the environment, such as material properties. Some of the questions we aim to answer in this article are:

- What is the performance of different calibrated propagation models—ranging from purely empirical to highly deterministic—in urban, suburban, and rural environments? In particular, can tuned empirical models perform well in complex urban environments? And, at the other extreme, are deterministic models overkill in rural environments?
- When models tuned in specific areas are to be applied to other areas that have no calibration data, what are the similarity features that matter the most in comparing areas to determine the best model for cross-application?
- What is the benefit of tuning and does the benefit vary per model category? In particular, do highly deterministic models reap any benefit from calibration and, if so, in which environments?

To answer these questions, we used more than 3 million samples of measurement data from five U.S. regions of varying size, population density, and terrain to provide a comprehensive evaluation of the most popular categories of pathloss models. To our knowledge, the breadth and depth of the data gathered are unprecedented in published material on this topic.

MEASUREMENT CAMPAIGN

In this section, we describe the measurement campaign conducted to gather the data samples used for analysis in this study. The data were furnished by the Infovista Corporation [11]. (Certain commercial equipment, instruments, or materials are identified in this article to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of

Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.)

MEASUREMENT DATA

Measurements to calibrate pathloss models for macrocell deployment most typically come in the form of signal-strength readings collected at a mobile device from neighboring base stations. More specifically, the signal strength from the station sectors is sampled during normal cellular operation and pinned to the GPS location of the mobile. Base stations usually have three sectors, each equipped with a directional antenna covering 120° in azimuth for a combined omnidirectional field of view.

Every technology will have its own metric for signal strength. For example, LTE uses reference signal receiver power (RSRP) while code-division multiple access (CDMA) uses received signal strength indicator (RSSI). The receiver device used for the measurement campaign was the PCTEL SeeGull MX Scanning Receiver. Its relevant manufacturer specifications are:

- *Power accuracy:* <1 dB above 50 °F
- *Receiver sensitivity:* -135 dBm at a 5-kHz bandwidth
- *Location accuracy:* <2.5 m
- *Heading accuracy:* <0.5°
- *Velocity accuracy:* <0.1 m/s.

The measurements were gathered by mounting the receiver antenna to the roof of a vehicle; for this reason, the campaign is referred to as a *drive test*. To convert the signal strength to pathloss, prior information associated with the sectors is necessary, namely, their physical cell identity (for sector

identification), transmit power, antenna pattern, and GPS location. Analogous prior information at the mobile station is also necessary, namely, the gain of the omnidirectional antenna, the amplifier gain, and the noise figure as well as any other components of the link budget, such as cable loss. The signal strength metric was converted to pathloss by deembedding the antenna patterns and the other losses so that the pathloss would reflect the channel alone and not the measurement equipment involved.

REGION DEFINITION

The five metropolitan regions considered in our study were Houston, New York, San Francisco, Boston, and Austin. As the vehicle moved about the regions during the drive tests, it fell into different sector-reception areas; we refer to them simply as *areas*. As an example, Figure 1 displays the color-coded signal strength for the 6,421 samples collected from a sector in Manhattan.

From the drive-test data collected, we down-selected 20 areas from each of the five regions for inclusion in this analysis. Down-selection of these 100 areas was imperative to reduce the data to a size at once manageable for processing yet large enough to sustain broad representation. The criterion for selection was an equal distribution between the three environments considered: urban, suburban, and rural. Our analysis over the 100 areas comprised more than 3 million data samples. Key properties of the base station antenna used in drive tests appear in Table 1 per region: the minimum, maximum, and mean values of their height, equivalent isotropically radiated power (EIRP), and azimuth and elevation beamwidths. Table 1 also catalogs the properties of the data samples per region: the minimum, maximum, and mean values of their RSRP (or RSSI for CDMA), base-mobile distances, and number of samples per area. In most regions, the operating technology was LTE in the 800-MHz band, with an exception in New York, for which it was CDMA at 1.5 GHz.

GEODATA

Geodata are provided in layers on a spherical grid indexed by latitude and longitude. Here, we employed the four most popular layers of geodata available, described in the sequel:

- *Terrain:* a numerical layer that specifies the elevation of the terrain above sea level at each grid point
- *Clutter:* a numerical layer that specifies the height of the clutter—skyscrapers, houses, trees, and so forth—above the terrain at each grid point
- *Usage:* a classification layer that indexes land usage at each grid point to a set of predefined classes. The usage layer available to us has 16 classes, listed in Table 2. We took advantage of it to partition the classes into urban, suburban, and rural environments, based on our own definition. Table 2 summarizes the partitioning. An area was then binned into one of the three environments according to which one occupied the majority of the area.



FIGURE 1. The signal-strength samples collected from a sector in the borough of Manhattan, New York City, during a drive test. The signal strengths are color coded against the legend.

TABLE 1. THE PROPERTIES OF THE MEASUREMENT CAMPAIGN.

Region		Houston	New York	San Francisco	Boston	Austin	
Base Station Antenna	Height (m)	Minimum	21	22.8	6.1	12.2	19.8
		Maximum	58.5	42.7	33.5	48.9	78
		Mean	40.4	30.2	18.5	27.3	36.6
	EIRP (dBm)	Minimum	56.5	46.5	55.6	56	60.3
		Maximum	58.1	57	57.7	57.9	60.3
		Mean	57.5	55	57.4	57.5	60.3
	Azimuth beamwidth (°)	Minimum	61	45.7	66.5	67.8	67.2
		Maximum	79.2	360	73.5	79.7	67.2
		Mean	71	101.1	68.5	69.3	67.2
	Elevation beamwidth (°)	Minimum	10.5	10.7	11.1	10.5	10.6
		Maximum	12.2	14	12.1	18.6	10.6
		Mean	11.4	11.3	11.8	13	10.6
Data Sample	RSSI (dBm)	Minimum	-135	-119	-135	-135	-133.5
		Maximum	-50.3	-29	-41.2	-48.6	-41.4
		Mean	-95.4	-93.9	-104.5	-98.1	-96.9
	Base-mobile distance (km)	Minimum	0.1	0.1	0.1	0.1	0.1
		Maximum	35	3.3	19	16	22.8
		Mean	3.8	0.9	3.2	2	3.7
	Number of samples per area (thousands)	Minimum	10.5	1	6.2	0.8	6.5
		Maximum	213.2	7.4	58.1	68	113
		Mean	70.3	4.2	20.1	14.4	47.5

■ *Morphology*: a set of polygons that delineates the complete 3D shape of the clutter beyond simply the clutter height.

Terrain and usage layers at a 30-m resolution are accessible free of charge from the United States Geological Survey (USGS). Clutter and morphology layers are accessible from providers per payment. The price of the geodata will depend on the resolution. In urban environments, geodata are available up to a 1-m resolution but, because of the cost, are usually purchased at a 5-m resolution, which is deemed sufficient for commercial accuracy. Clutter is available in suburban environments from a 10-m to a 15-m resolution; morphology is rare to find, except in urban environments, because there is little demand. Here, we used a 5-m resolution in the urban environment, 15 m in the suburban, and 30 m in the rural.

PATHLOSS MODELS

In this section, we describe the six pathloss models that were investigated in our study. They span a broad range of categories and are introduced in order from the purely empirical to the increasingly deterministic. While there are many models available per category ([13] and [14] provide excellent surveys), we have selected the most widely employed throughout the propagation community in the macrocell bands (700–2,100 MHz).

HATA MODEL

The original Hata model [15], often referred to as the *Okumura-Hata model*, was derived empirically from measurements collected in urban Tokyo in the 1960s. It is limited to below 1.5 GHz by the range of frequencies over which the data were gathered. Its intrinsic parameters are the coefficients associated

TABLE 2. THE CLASSES IN THE USAGE LAYER.

Urban Environment	Suburban Environment	Rural Environment
Urban	Open	Rural
Core urban	Airport	Seawater
High-density urban	Transportation	Inland water
Open in urban	Commercial industrial	Marsh wetland
	Residential with trees	Grassland agriculture
	Residential with few trees	Forested dense vegetation

with the model variables, namely, the center frequency, base-mobile distance, and the base and mobile antenna heights. Because the coefficients quantify the degree of variation, or slope, with respect to each variable, this model category is often referred to as *slope based* [16], [17]. Because of simplicity and, in turn, computational efficiency, it is very popular [18]–[20]. The only layer utilized by the model is the usage layer, through which the pathloss constant is tuned per usage class to improve the fit to the measurement data [1]. The constants found can range up to 30 dB or higher; hence, they are key to fitting the model predictions to the measurements.

The Hata model has since expanded from the original to versions for suburban and rural environments as well, each with a distinct set of coefficients [21]. Note that, in our study, the coefficients of the Hata model were not tuned, in contrast to [8]–[10], because the model would no longer be consistent with the one that (despite its poor performance, as we will see) is still widely employed today.

There are also more complex slope-model variations specific to urban environments. The Bertoni–Walfisch model [2], for example, reduces propagation around buildings to multiple screen diffractions while the Walfisch–Ikegami model [3] introduces supplementary variables for the height and separation of buildings and for the width and orientation of roads.

CLOSE-IN MODEL

There are also generic slope-based models whose coefficients are derived exclusively through calibration [4]–[6]. One such model widely used today is the close-in model [7], which has only two parameters: 1) a tunable coefficient for the base-mobile distance and 2) a pathloss constant set to the free-space propagation loss at 1 m, a physical constraint. The benefit of the model is its simplicity: it uses no geodata, and the single tunable parameter facilitates the comparison of propagation in different environments. For the untuned version, the coefficient is simply set to two, equating it to the free-space propagation model.

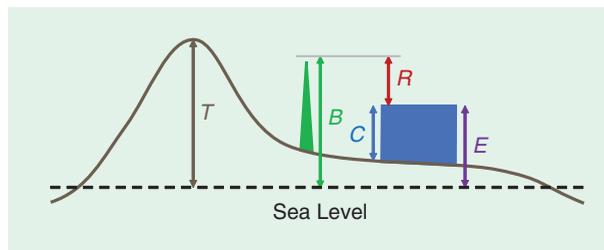


FIGURE 2. An illustration of the five proposed height features: *T*: terrain; *C*: clutter; *E*: effective; *B*: base; *R*: relative.

FLOATING-INTERCEPT MODEL

The floating-intercept model [7] is almost identical to the close-in model, with the exception that the physical constraint is relaxed so that the pathloss constant can be tuned to deliver a better fit, at the price of slightly higher complexity. The same free-space propagation model as for the close-in model is used for the untuned version.

Although the prediction error was generally highest in the urban environment, it was there that calibration was most beneficial.

LONGLEY–RICE MODEL

The Longley–Rice model [22], often referred to as the *irregular terrain model*, was the first in its category of terrain models [23], dating back to 1982, and is still the most widely employed.

It is appropriately named, for its prediction is based on the terrain layer. Specifically, it treats the two or three points on the terrain profile that most obstruct the first Fresnel zone [24] between the base and mobile stations as knife edges. A marching algorithm is implemented to propagate the field strength between successive obstruction points through the two-path model: 1) the diffracted path from the knife edge computed from the Fresnel–Kirchoff formula plus 2) the path from the ground reflection. The stochasticity of the model is embodied in the attenuation of the paths to adjust for situation, time, and location variability. The curvature of Earth and tropospheric scattering are also accounted for. As with the Hata model, the Longley–Rice model calibrates the pathloss constant against the measurement data per usage class.

COMMUNICATIONS RESEARCH CENTRE PREDICT MODEL

The Communications Research Centre (CRC) predict model [25] is, in principle, similar to the Longley–Rice but bears significant differences, making it more accurate, albeit more computationally intensive. The CRC predict model takes advantage of both the terrain and clutter layers—hence falling into the category of clutter models [26]—to account for the effective diffraction height by computing a composite profile from the two. All obstruction points over the composite profile within the first Fresnel zone are considered, not just the most dominant. This is critical for urban environments in which there are numerous knife edges. At each obstruction, the field strength at multiple, variably spaced points above the knife edge are considered to render a more exact description of the wave field. A marching algorithm propagates the field strength at the points between consecutive obstructions using the same two-path model. Given the clutter layer, for enhanced precision, the calibration process allows for tuning of the clutter height and clutter separation per usage class in addition to the constant.

ORANGE LABS 3D RAY-TRACING MODEL

The most deterministic of the categories examined in our study are 3D ray-tracing models [27], [28]. The one designed

by Orange Labs is included in such popular cell planning tools as Mentum Planet [29] and Atoll [30] and, as such, was chosen to represent this category here. In addition to diffraction and reflection in the vertical plane, the two are accounted for in the horizontal plane as well, made possible solely through the morphology layer. Describing the clutter through closed polygons also enables distinguishing between indoor and outdoor environments, to which different pathloss constants can be assigned when such measurement data are available. Thus, the model exploits all four geodata layers, for both prediction and tuning. Finally, the model incorporates reflections from natural landscape features, such as hills and mountains. This model category was prohibitive in the past because of the computational intensity required but has become popular in recent years, thanks to faster processors.

For convenience, we denote the set of models as $\mathcal{M} = \{H/CI/FI/LR/P/3D\}$, where H indicates the Hata model, CI the close-in model, FI the floating-intercept model, LR the Longley–Rice model, P the CRC predict model, and $3D$ the Orange Labs 3D ray-tracing model.

MODEL CROSS-APPLICATION

In this section, we propose a methodology to select, among a candidate set of calibrated models, the one most suitable for application to an area with no measurement data. The selection process is based on feature similarity between the calibrated areas and the application area, what we refer to pairwise as *cross-application areas*. As far as we could find, there is just one reference in the open literature that treats this issue [31]. The reference, however, is limited to the Hata model and includes only 10 areas in a single urban region in the analysis. Moreover, while the reference proposes features to determine similarity, it does not reveal which are the most effective. In a related work [32], a neural network is trained on data sets in one area using area features as inputs and then applied to a different area to evaluate how well it can generalize. The study includes just three areas in a rural environment.

AREA FEATURES AND SIMILARITY METRICS

The similarity between two areas is based on area features, also referred to as *signatures* or *fingerprints* because they serve as proxies for the areas. In this work, we propose six features, five of which are based on area height: terrain (T), clutter (C), effective (E), base (B), and relative (R), as depicted in Figure 2. The elevation of the terrain and the height of the clutter above the terrain are obvious features to characterize similarity and so are taken directly from the respective geodata layers. A composite layer we define as the effective height $E = T + C$ is entertained because the vertical diffraction height for models that take clutter into account is based on the combined height of the two. The height of the base station is important because results will vary particularly on whether the antenna is above or below the clutter, potentially generating very different results. In fact, to target this feature explicitly, another composite layer,

the relative height R between the base and effective heights $R = B - E$, is also entertained. The latter is the sole height feature that can be negative.

The five height features are all computed through some addition (or subtraction) of the terrain and clutter layers (which are numerical grids) and the height of the base station. Since the features are all based on numerical values, the similarity between two areas is, in turn, numerical. Specifically, let each grid point \mathbf{x}_i^n be indexed through $n = 1 \dots N_i$, where N_i is the number of samples in area i . Furthermore, let the grid point have associated feature value y_i^n (T , C , E , B , or R) from which that cumulative distribution function $F_i(y)$ characterizing the area is derived. The similarity metric between areas i and j is then defined as:

$$s_{ij}^{T/C/E/B/R} = 1 - \max_y |F_i^{T/C/E/B/R}(y) - F_j^{T/C/E/B/R}(y)|. \quad (1)$$

The metric is derived from the Kolmogorov–Smirnov goodness-of-fit test [33]: the metric will be one if the areas have exactly the same distribution and zero if they have no features in common. (The Kullback–Leibler divergence [33] was also considered as an alternative to the Kolmogorov–Smirnov goodness-of-fit test. The feature weights between the former and the latter varied by 4.7% at most.)

The sixth feature is based on the usage layer. Recall that, in contrast to the terrain and clutter layers, it is not numerical; rather, each grid point indexes one of the 16 usage classes. It follows that the distribution of usage in area i is expressed as the fraction (a number between zero and one) of the total area assigned to each class and is recorded as a 16-element row vector \mathbf{v}_i . As the distribution is expressed as a discrete vector, it necessitates an equivalent discrete similarity metric. Accordingly, the similarity metric between areas i and j is calculated as

$$s_{ij}^U = \frac{\mathbf{v}_i \cdot \mathbf{v}_j^t}{\sqrt{(\mathbf{v}_i \cdot \mathbf{v}_i^t) \cdot (\mathbf{v}_j \cdot \mathbf{v}_j^t)}}, \quad (2)$$

where t denotes the transpose operation and U indicates usage. The value is one if $\mathbf{v}_i = \mathbf{v}_j$ and zero if the areas have no usages in common.

For convenience, we introduce the similarity vector $\mathbf{s}_{ij} = [s_{ij}^T \ s_{ij}^C \ s_{ij}^E \ s_{ij}^B \ s_{ij}^R \ s_{ij}^U]$, which contains the similarity metrics for all six features. Note that $\mathbf{s}_{ii} = \mathbf{1}$, $\forall i$ and that the similarity metric is commutative, i.e., $\mathbf{s}_{ji} = \mathbf{s}_{ij}$. Also, for convenience, we denote the feature set as $\mathcal{F} = \{T/C/E/B/R/U\}$.

CROSS-APPLICATION ERROR AND CROSS-APPLICATION GAIN

We define the cross-application error when applying a model tuned on area j to area i as:

$$\epsilon_{ij}^M = \sqrt{\frac{1}{N_i} \sum_{n=1}^{N_i} [PL(\mathbf{x}_i^n) - PL_j^M(\mathbf{x}_i^n)]^2}, \quad (3)$$

where $PL(\mathbf{x}_i^n)$ is the measured pathloss at grid point \mathbf{x}_i^n (from the GPS location of the mobile) and $PL_j^M(\mathbf{x}_i^n)$ is the pathloss predicted by model M tuned on area j . The predictions were generated by Mentum Planet.

To gauge the benefit, or gain, of applying a calibrated model to area i compared to the model with nominal parameters, we define the cross-application gain as

$$\Delta \epsilon_{ij}^M = \epsilon_{i,\text{UNCAL}}^M - \epsilon_{ij}^M, \quad (4)$$

where $\epsilon_{i,\text{UNCAL}}^M$ denotes the error from the uncalibrated model. The value is positive if tuning improves the performance of the model and negative otherwise. Note that the cross-application error and cross-application gain are not commutative, i.e., $\epsilon_{ij}^M \neq \epsilon_{ji}^M$, $\Delta \epsilon_{ij}^M \neq \Delta \epsilon_{ji}^M$.

FEATURE WEIGHTS

Part of the methodology to select the most suitable calibrated model is determining which of the proposed features is the most effective in estimating the cross-application error from the similarity metric. A basic criterion is that high similarity between two areas should correlate with low error. Accordingly, we gauge the effectiveness of feature \mathcal{F} casewise—a case being defined by model \mathcal{M} in a given environment—through the Pearson correlation coefficient [33] between $s_{ij}^{\mathcal{F}}$ and ϵ_{ij}^M , expressed as

$$\rho_{s_{ij}^{\mathcal{F}}, \epsilon_{ij}^M} = \frac{\sum_{i=1}^P \sum_{j=1}^P (s_{ij}^{\mathcal{F}} - \mu_{s_{ij}^{\mathcal{F}}}) \cdot (\epsilon_{ij}^M - \mu_{\epsilon_{ij}^M})}{\sqrt{\sum_{i=1}^P \sum_{j=1}^P (s_{ij}^{\mathcal{F}} - \mu_{s_{ij}^{\mathcal{F}}})^2} \sqrt{\sum_{i=1}^P \sum_{j=1}^P (\epsilon_{ij}^M - \mu_{\epsilon_{ij}^M})^2}}, \quad (5)$$

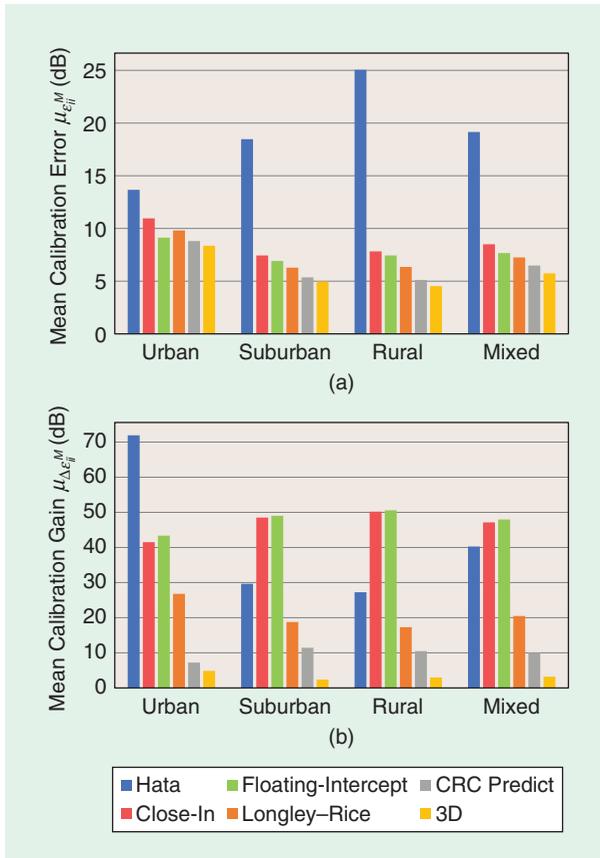


FIGURE 3. (a) The mean calibration error and (b) mean calibration gain for the 16 cases investigated.

where P denotes the number of areas in the environment and μ the mean of the said quantities across the P^2 possible cross-application areas. The feature with the most negative coefficient (since the two quantities should be inversely correlated) is identified as the most effective feature.

In our experience, we have found that joint features are typically more effective than any single feature alone. To that end, we introduce the weight vector $\mathbf{w} = [w^T w^C w^E w^B w^R w^U]$ to represent the features collectively. The vector space is then searched to minimize the correlation coefficient between the weighted similarity metric $s_{ij}^w = \mathbf{w} \cdot \mathbf{s}_{ij}^t$, and the cross-application error. The problem can be stated precisely as a nonlinear program:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \rho_{s_{ij}^w, \epsilon_{ij}^M} \\ \text{s.t.} \quad & \|\mathbf{w}\|_1 = 1. \\ & \mathbf{w} > \mathbf{0} \end{aligned} \quad (6)$$

The solution to the program was found exhaustively with a search granularity of 0.05 for each weight, yielding the optimal weight vector $\hat{\mathbf{w}}$. The resultant optimal weighted similarity metric s_{ij}^w serves to identify the candidate area j that is most similar to application area i for the case. In turn, the candidate model tuned on area j is then selected for application.

RESULTS

In this section, we present the results for the six models listed in the “Pathloss Models” section versus four environments—urban, suburban, rural, and mixed (a combination of the three)—for a total of 16 cases.

CALIBRATION ERROR AND CALIBRATION GAIN

The primary objective of the calibration process is to minimize the calibration error, i.e., the root-mean-square error between predictions from model \mathcal{M} and the measurement samples in area i . Note that this simply corresponds to the cross-application error from area i to area i , ϵ_{ii}^M . The mean calibration error over all areas in the case environment is reported in Figure 3(a) for the 16 cases. We also analyze the calibration gain, i.e., the reduction in calibration error due to model tuning. Analogously, this corresponds to the cross-application gain $\Delta \epsilon_{ii}^M$. Figure 3(b) reports the mean calibration gain.

The urban environment is the most challenging because of the varying landscape due to the tall buildings in the city center that cause more uncertainty in propagation; in fact, we observe the highest calibration error there. The value for the close-in model is $\mu_{\epsilon_{ii}^{ci}} = 11$ dB and drops to $\mu_{\epsilon_{ii}^{fi}} = 9.1$ dB for the floating-intercept model. This is impressive, considering that neither model exploits any geodata. Nevertheless, the three deterministic models deliver mostly better results. The value for Longley–Rice is $\mu_{\epsilon_{ii}^{lr}} = 9.8$ dB. Recall that this model just accounts for terrain and usage, so it has no perception of the clutter on the horizon. It can then only adjust the pathloss constant per usage class based on the greater loss experienced in the city center. By exploiting supplemental data about the vertical height of the structures accessible through the clutter layer, the CRC predict model is able to decrease the uncertainty, improving the error to

$\mu_{\epsilon_{ii}^e} = 8.8$ dB. The 3D model decreased the error yet further to $\mu_{\epsilon_{ii}^{3D}} = 8.3$ dB by taking advantage of the morphology layer to also capture the horizontal propagation.

Due to the simpler landscape in the suburban environment, thanks primarily to residential housing, the error of the close-in model decreased to $\mu_{\epsilon_{ii}^{CI}} = 7.4$ dB, a significant amount with respect to the urban environment. The other models followed suit, maintaining their relative performance: $\mu_{\epsilon_{ii}^{FI}} = 6.9$ dB, $\mu_{\epsilon_{ii}^{LR}} = 6.3$ dB, $\mu_{\epsilon_{ii}^p} = 5.4$ dB, and $\mu_{\epsilon_{ii}^{3D}} = 5$ dB. For the rural environment, the calibration errors were generally lower compared to the suburban environment (essentially due to minimal clutter impeding the radio path), and the same trend across the models was sustained: $\mu_{\epsilon_{ii}^{CI}} = 7.8$ dB, $\mu_{\epsilon_{ii}^{FI}} = 7.4$ dB, $\mu_{\epsilon_{ii}^{LR}} = 6.3$ dB, $\mu_{\epsilon_{ii}^p} = 5.1$ dB, and $\mu_{\epsilon_{ii}^{3D}} = 4.5$ dB. And when considering the overall performance in the mixed environment, the trend was also consistent: $\mu_{\epsilon_{ii}^{CI}} = 8.5$ dB, $\mu_{\epsilon_{ii}^{FI}} = 7.7$ dB, $\mu_{\epsilon_{ii}^{LR}} = 7.2$ dB, $\mu_{\epsilon_{ii}^p} = 6.5$ dB, and $\mu_{\epsilon_{ii}^{3D}} = 5.7$ dB.

The trends for the five models commented upon thus far were not sustained for the Hata model. The calibration error for the urban environment was actually lowest ($\mu_{\epsilon_{ii}^H} = 13.6$ dB) and then increased for the suburban ($\mu_{\epsilon_{ii}^H} = 18.4$ dB) and rural ($\mu_{\epsilon_{ii}^H} = 25$ dB) environments. Recall that the Hata model is purely empirical, meaning that its coefficients were derived from measurements taken in areas different—in most cases, extremely different—from the ones investigated here. As such, the trend suggests that, when comparing the areas investigated versus the ones where the empirical measurements for the Hata model were taken, the areas in the urban environment were more similar than the ones in the suburban and rural environments in terms of model coefficients, the parameters that were not tuned.

Although the calibration error was generally highest in the urban environment across the models, it was there that the highest calibration gain was witnessed. The gain diminished in the suburban environment and fell yet further in the rural environment, meaning that tuning was most beneficial in complex environments. Between the six models, tuning clearly benefited the slope-based models the most, with a gain of $\mu_{\Delta\epsilon_{ii}^H} = 40.3$ dB, $\mu_{\Delta\epsilon_{ii}^{CI}} = 47.2$ dB, and $\mu_{\Delta\epsilon_{ii}^{LR}} = 47.9$ dB in the mixed environment. That value dropped to $\mu_{\Delta\epsilon_{ii}^{LR}} = 20.5$ dB for Longley–Rice, to $\mu_{\Delta\epsilon_{ii}^p} = 10$ dB for the CRC predict model, and down to $\mu_{\Delta\epsilon_{ii}^{3D}} = 3.2$ dB for the 3D model. This indicates that the benefit of calibration diminished as the model became more deterministic.

As in our study, calibration errors reported in previous works vary widely depending on the model, the environment, and the specific areas measured: 5.4–37.1 dB for the urban environment [2], [4], [8], [16], [17], [20], [21], [24], 3.5–37.7 dB for the suburban [5], [16], [21], and 5–43.9 dB for the rural [16], [20], [21].

FEATURE WEIGHTS

Figure 4(a) displays the cross-application error ϵ_{ij} versus the optimal similarity metric s_{ij}^w for the Longley–Rice model in the mixed environment as an illustrative case. The optimal weight vector $\hat{\mathbf{w}}$ resulted from the solution to (6). Each point in the plot indexes area i crossed with area j , for a

total of $P^2 = 10,000$ unique combinations ($P = 100$ in the mixed environment, comprising 20 areas in each of the five regions). The moving average of the points over the similarity is shown in red. It is obvious from the plot that the average error decreases with increasing similarity, reflecting that the two are negatively correlated; indeed, the optimal correlation coefficient is $\rho_{s_{ij}^w, \epsilon_{ij}^{LR}} = -0.60$. Note that $s_{ij}^w = 1$ indicates that the areas are identical ($j = i$) and so the average cross-application error at this similarity value simply corresponds to the mean calibration error, highlighted in the figure. Also shown, in magenta, is the moving average plus and minus the moving standard deviation. It also decreases with similarity, meaning that uncertainty in the prediction error also decreases as the areas become more similar.

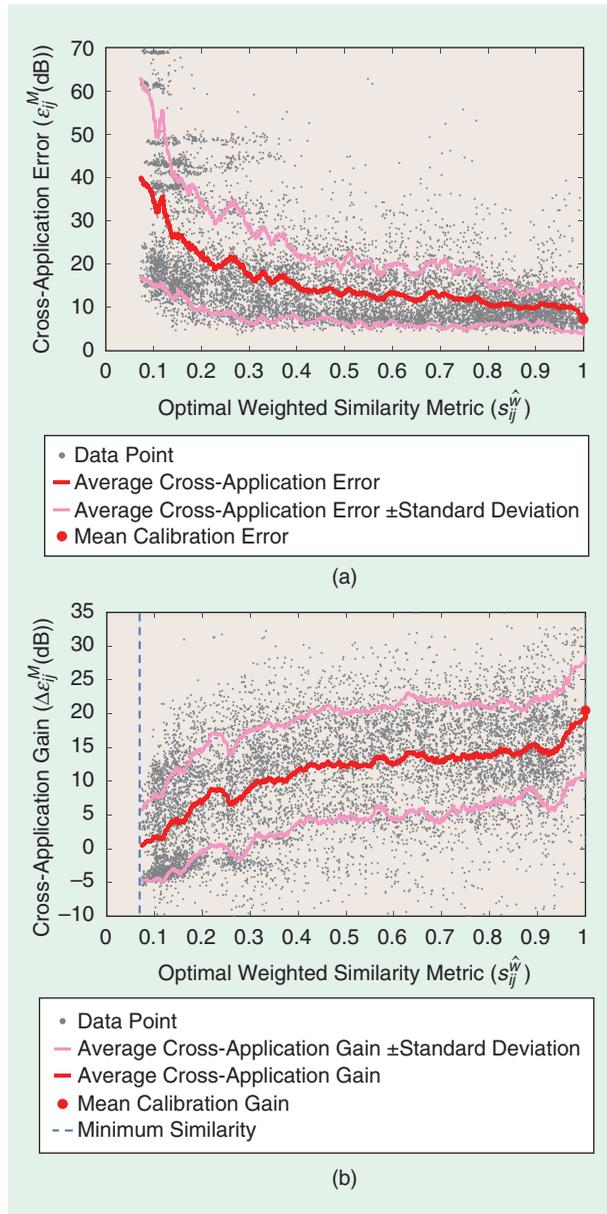


FIGURE 4. (a) The cross-application error and (b) cross-application gain for an illustrative case of the Longley–Rice model in the mixed environment.

We now report the optimal feature weights \hat{w} for all 16 cases. Figure 5(a)–(c) shows the weights for the clutter, base, and usage features only. The weights for the terrain, effective, and relative features were omitted because they all averaged less than 0.02 across the cases. The terrain was mostly flat compared to the clutter (the only exception being in San Francisco); hence, it was found to be negligible. Since the effective and relative heights are composite layers of the clutter, base, and terrain layers, their marginal weights suggest that they are redundant and can otherwise be accounted for directly through those layers.

When examining the urban environment, the clutter weight in Figure 5(a) was the most dominant ($\hat{w}^{C/CI} = 0.6$, $\hat{w}^{C/Fl} = 0.5$, $\hat{w}^{C/LR} = 0.65$, $\hat{w}^{C/P} = 0.5$, $\hat{w}^{C/3D} = 0.55$), indicating that clutter is the best feature to identify similarity there. This makes intuitive sense since, in cities, prediction will depend mostly on the obstructive clutter. The base feature

was also significant for the deterministic models ($\hat{w}^{B/LR} = 0.2$, $\hat{w}^{B/P} = 0.4$, $\hat{w}^{B/3D} = 0.35$)—important in distinguishing whether transmission occurred above or below the clutter. Base height is not accounted for in the close-in and floating-intercept models. Accordingly, it makes sense that its feature weights are both close to zero. In fact, this is true for all environments, though it is shown in [34] that the close-in model can effectively account for base height when it is introduced as an additional variable.

For the rural environment, because clutter height is negligible, its feature weights, in turn, were also negligible across the board. The base feature, however, as in the urban environment, was significant for the deterministic models since they account for it ($\hat{w}^{B/LR} = 0.8$, $\hat{w}^{B/P} = 1$, $\hat{w}^{B/3D} = 0.75$). Since the suburban environment lies somewhere between the urban and rural, it is not surprising that the optimal weights were inconclusive across the models there.

When analyzing all of the environments together, the usage feature was the most dominant. Its weight was very high for the Longley–Rice and CRC predict models ($\hat{w}^{U/LR} = 0.85$ and $\hat{w}^{U/P} = 0.9$). Recall what was mentioned in the “Hata Model” section, i.e., that a key parameter for tuning is the pathloss constant per usage class. The values reported then suggest that the Longley–Rice and CRC predict models rely heavily on tuning. The values are also high for the close-in and floating-intercept models but to a lesser degree since their constants are not usage specific. The usage weight for the mixed environment dropped to $\hat{w}^{U/3D} = 0.15$ for the 3D model, suggesting that it relies much less on tuning compared to the other models. This is also not surprising, given the highly deterministic nature of the 3D model. In fact, for the 3D model, it is the base feature that prevailed: $\hat{w}^{U/3D} \geq 0.7$ for three out of the four environments. But the base layer is not area dependent at all; it does not quantify how similar two areas were but rather how similar the base heights during calibration were, another indicator that the 3D model is less dependent on area tuning.

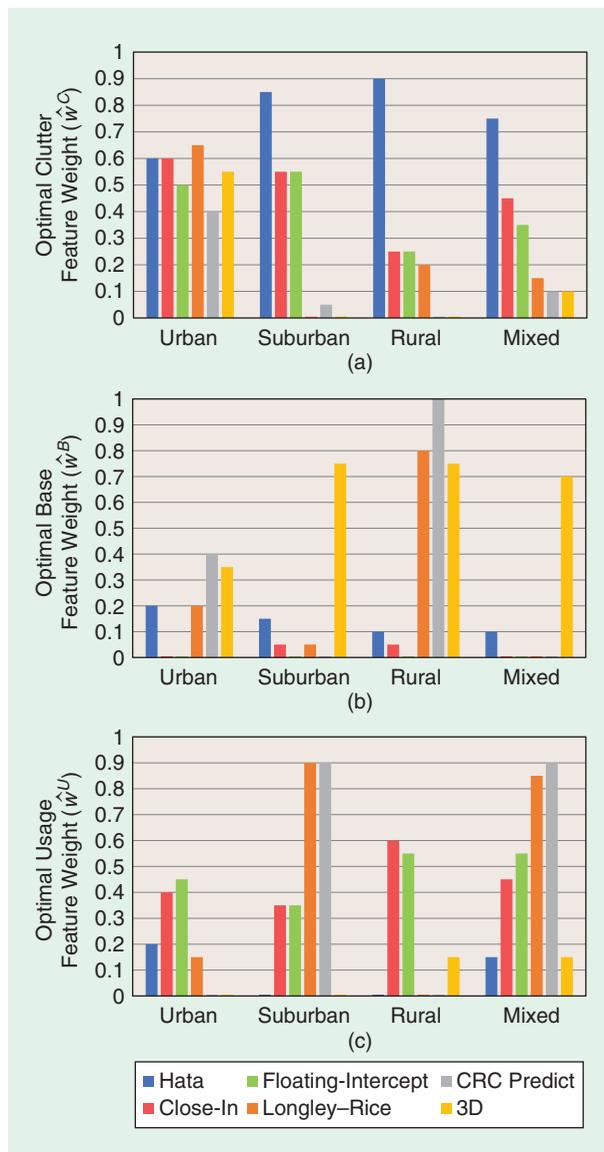


FIGURE 5. The optimal (a) clutter, (b) base, and (c) usage feature weights for the 16 cases investigated.

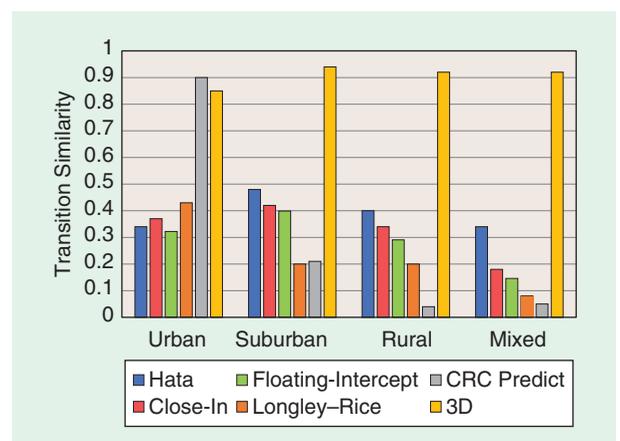


FIGURE 6. The transition similarity for the 16 cases investigated. The framed bars signify that the average cross-application gain was positive for all similarities recorded for the case. Hence, the transition similarity simply defaulted to the minimum similarity recorded.

There is another reason that usage is the most dominant in the mixed environment. Recall from the “Geodata” section that the usage layer was exploited to partition all of the areas investigated into urban, suburban, and rural environments by usage class. Thus, one would expect that the usage feature weight within an individual environment would be low, i.e., it might not be a strong discriminator between areas within the same environment because the areas are already somewhat similar. And the usage feature will certainly be a better discriminator when the three environments are mixed together.

Yet another indicator of model dependency on tuning is the correlation coefficient $\rho_{s_{ij}^w, \epsilon_{ij}^M}$. As seen earlier, $\rho_{s_{ij}^w, \epsilon_{ij}^{LR}} = -0.6$ for the case of Longley–Rice in the mixed environment. The absolute value of the coefficient is relatively large, attesting that the similarity metric is well correlated with the prediction error; in other words, if the areas are similar, cross-application can significantly reduce the cross-application error. The corresponding values for the close-in, floating-intercept, and CRC predict models were $\rho_{s_{ij}^w, \epsilon_{ij}^{CI}} = -0.33$, $\rho_{s_{ij}^w, \epsilon_{ij}^{FI}} = -0.34$, and $\rho_{s_{ij}^w, \epsilon_{ij}^R} = -0.36$. On the other hand, the 3D model being the most deterministic of the given models, the prediction success of the model is least dependent on area similarity: $\rho_{s_{ij}^w, \epsilon_{ij}^{3D}} = -0.19$.

The model whose cross-application error is least correlated with the similarity metric is the Hata model: $\rho_{s_{ij}^w, \epsilon_{ij}^H} = -0.17$. Given the model behavior observed in the “Calibration Error and Calibration Gain” section—that there is worse performance in the rural environment than the urban environment and the rationale behind it—this is not unexpected. While the correlation coefficient gauges the relative benefit of cross-application when the areas are similar, it does not indicate the benefit with respect to the alternative to cross-application, i.e., applying the uncalibrated model; that is what we probe next.

CROSS-APPLICATION GAIN

To ascertain the actual benefit of cross-application, we examine the cross-application gain. Figure 4(b) displays $\Delta\epsilon_{ij}^M$ versus s_{ij}^w , again for the illustrative case of Longley–Rice in the mixed environment. Also displayed are the moving average and the plus and minus moving standard deviation. Recall that a positive gain ($\Delta\epsilon_{ij}^M > 0$) signifies that the model tuned on area j is more beneficial than the uncalibrated model. Since the average is positive over the full range of the similarities recorded, it is expected that any tuned model will be more beneficial for this case; the expectation improves with increasing similarity, from a 0.5-dB gain at $s_{ij}^w = 0.07$ to a 20.5-dB gain at $s_{ij}^w = 1$. As before, $s_{ij}^w = 1$ indicates that the areas are identical, and so the average cross-application gain at this similarity value simply corresponds to the mean calibration gain, highlighted in the figure.

The deterministic models outperformed the empirical models not just in the urban environment, as expected, but in the rural environment as well.

To identify when cross-application was beneficial over all 16 cases, each was examined separately to determine beyond which similarity the average cross-application gain sustained only positive values; we refer to this value as the *transition similarity*, a final indicator of model dependency on tuning. Figure 6 displays the results. The framed bars signify that the average gain was reported positive for the full range of similarities (as for the previous illustrative case), and

so the transition similarity simply defaulted to the minimum similarity recorded for the case [see Figure 4(b)]. This was true for the Hata, close-in, floating-intercept, and Longley–Rice models in all environments and for the CRC predict model with an exception in the urban environment. Only for the 3D model was there a genuine transition in all environments, meaning that it was beneficial to use a calibrated model only for very high similarity ($s_{ij}^w \geq 0.85$), and that the model otherwise predicted reliably even without tuning.

Note that, when examining the framed bars, the minimum similarity recorded was much smaller, in general, for the mixed environment than for the separate environments. This is because binning the areas into the three environments augmented the similarity within each.

CONCLUSIONS

In this article, we conducted a comprehensive analysis of calibrated pathloss models using over 3 million measurement samples from five metropolitan regions throughout the United States while considering six pathloss models, ranging from the purely empirical close-in model to the highly deterministic 3D model.

In the urban environment—the most challenging of the three environments—the prediction error was 11 dB for the close-in model and dropped to 8.3 dB for the 3D. The error decreased across the models in the suburban environment (7.4 dB for close-in and 5 dB for 3D) and yet further in the rural environment for most models (7.8 dB for close-in and 4.5 dB for 3D). Hence, the deterministic models outperformed the empirical models not just in the urban environment, as expected, but in the rural environment as well. Although the prediction error was generally highest in the urban environment, it was there that calibration was most beneficial, reducing the error up to 41.5 dB for the close-in model to as low as 4.5 dB for the most deterministic model. The performance of the Hata model was much worse than the other five investigated.

Regarding model cross-application, in identifying similarity between areas, we found that the clutter feature was the most effective in the urban environment, the base feature in the rural area, and the usage feature in the mixed environment while the results varied in the suburban environment.

Finally, five of the six models benefited from cross-application across all environments, regardless of similarity. The only exception was the 3D model, which benefited only when the similarity was very high (above 0.85 across all environments), meaning that it was able to predict reliably even without calibration.

ACKNOWLEDGMENTS

We would like to thank the Infovista Corporation and in particular its employees Jason Suplita and Darren Hart for having furnished the invaluable drive-test data, essential to the analysis and results presented in this article.

AUTHOR INFORMATION

Jiayi Zhang (jiayi.zhang@nist.gov) is a contractor at the National Institute of Standards and Technology, Gaithersburg, Maryland. He received his Ph.D. degree in electrical engineering from the University of Southampton, United Kingdom, in 2012.

Camillo Gentile (camillo.gentile@nist.gov) is a project leader in the Wireless Networks Division of the National Institute of Standards and Technology, Gaithersburg, Maryland. He received his Ph.D. degree in electrical engineering from Pennsylvania State University, University Park, in 2001.

Wesley Garey (wesley.garey@nist.gov) is a member of the Wireless Networks Division of the National Institute of Standards and Technology, Gaithersburg, Maryland. He received his M.S. degree in computer science from John Hopkins University, Baltimore, Maryland, in 2017.

REFERENCES

- [1] M. A. Alamoud and W. Schutz, "Okumura-Hata model tuning for TETRA mobile radio networks in Saudi Arabia," in *Proc. IEEE 2nd Int. Conf. Advances in Computational Tools for Engineering Applications (ACTEA)*, Dec. 2012, pp. 47–51.
- [2] B. Y. Hanci and I. H. Cavdar, "Mobile radio propagation measurements and tuning the path loss model in urban areas at GSM-900 band in Istanbul-Turkey," in *Proc. IEEE 60th Vehicle Technology Conf. (VTC-Fall)*, vol. 1, Los Angeles, CA: IEEE, Sept. 2004, pp. 139–143.
- [3] J. V. Rodriguez, S. G. Martinez, C. Garca-Pardo, J.-M. Molina-Garcia-Pardo, and L. Juan-Llaser, "Comparison of various urban radiowave propagation models with measurements," *IEEE Antennas Wireless Propag. Lett.*, vol. 8, pp. 977–980, Aug. 2009.
- [4] A. Mousa, Y. Dama, M. Najjar, and B. Alsayeh, "Optimizing outdoor propagation model based on measurements for multiple RF cell," *Int. J. Comput. Appl.*, vol. 60, no. 5, pp. 5–10, Jan. 2012.
- [5] J. Milanovic, S. Rimac-Drlje, and K. Bejuk, "Comparison of propagation models accuracy for WiMAX on 3.5 GHz," in *Proc. 14th IEEE Int. Conf. Electronics, Circuits, and Systems (ICECS)*, Dec. 2007, pp. 111–114.
- [6] M. N. Abdallah et al., "Further validation of an electromagnetic macro model for analysis of propagation path loss in cellular networks using measured driving-test data," *IEEE Antennas Propag. Mag.*, vol. 56, no. 4, pp. 108–129, Aug. 2014.
- [7] M. K. Samimi, T. S. Rappaport, and G. R. MacCartney, "Probabilistic omnidirectional path loss models for millimeter-wave outdoor communications," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 357–360, Aug. 2015.
- [8] R. Verma and G. Saini, "Statistical tuning of Cost-231 Hata model at 1.8 GHz over dense urban areas of Ghaziabad," in *Proc. 3rd IEEE Int. Conf. Computing for Sustainable Global Development*, Mar. 2016, pp. 1220–1225.
- [9] O. Banimelhem, M. M. Al-Zu'bi, M. S. Al Salameh, "Hata path loss model tuning for cellular networks in Irbid City," in *Proc. IEEE Int. Conf. Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM)*, Oct. 2015, pp. 1646–1650.

- [10] B. S. Castro, M. R. Pinheiro, G. P. dos Santos Cavalcante, I. R. Gomes, and O. De Oliveira Carneiro, "Comparison between known propagation models using least squares tuning algorithm on 5.8 GHz in Amazon region cities," *J. Microw. Optoelectron. Electromagn. Appl.*, vol. 10, no. 1, pp. 106–113, June 2011.
- [11] Infovista. Accessed on: Aug. 2019. [Online]. Available: www.infovista.com
- [12] L. Simić, J. Riihijärvi, and P. Mähönen, "Can statistical propagation models be saved by real 3D city data? A regionalized study of radio coverage in New York City," in *Proc. IEEE Dynamic Spectrum Access Networks (DySPAN)*, Sept. 2015, pp. 285–288.
- [13] C. Phillips, D. Sicker, and D. Grunwald, "A survey of wireless path loss prediction and coverage mapping methods," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 255–270, Mar. 2013.
- [14] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [15] Y. Okumura, "Field strength and its variability in VHF and UHF land-mobile radio service," *Rev. Electr. Commun. Lab.*, vol. 16, pp. 825–873, 1968.
- [16] M. S. Mollel and M. Kisangiri, "An overview of various propagation model for mobile communication," in *Proc. IEEE Science, Computing and Telecommunications*, July 2014, pp. 148–153.
- [17] D. Sharma, P. K. Sharma, V. Gupta, and R. K. Singh, "A survey on path loss models used in wireless communication system design," *Int. J. Recent Trends Eng. Technol.*, vol. 3, no. 2, pp. 115–118, May 2010.
- [18] G. Sati and S. Singh, "Comparison of various outdoor propagation models with empirical data collection using TEMS 10.0.5 for DIT university, Dehradun," *Int. J. Comput. Appl.*, vol. 97, no. 10, pp. 9–13, July 2014.
- [19] N. Faruk, Y. A. Adediran, and A. A. Ayeni, "Error bounds of empirical path loss models at VHF/UHF bands in Kwara State, Nigeria," in *Proc. IEEE Eurocon*, July 2013, pp. 602–607.
- [20] J. Baumgarten, K. Lien Chee, A. Hecker, T. Kürner, M. Braum, and P. Zahn, "Performance of prediction models in suburban/rural residential areas at 860, 2300 and 3500 MHz," in *Proc. IEEE 6th European Conf. Antennas and Propagation (EUCAP)*, Mar. 2012, pp. 1412–1416.
- [21] V. S. Abhayawardhana, I. J. Wassell, D. Crosby, M. P. Sellars, and M. G. Brown, "Comparison of empirical propagation path loss models for fixed wireless access systems," in *Proc. IEEE 61st Vehicle Technology Conf. (VTC-Spring)*, vol. 1, Stockholm, Sweden: IEEE, May 2005, pp. 73–77.
- [22] G. A. Hufford, A. G. Longley, and W. A. Kissick, "A guide to the use of the ITS irregular terrain model in the area prediction mode," U.S. Dept. Commerce NTIA, Washington, D.C., NTIA Rep. 82-100, Apr. 1982.
- [23] T. Kurner, D. J. Cichon, and W. Wiesbeck, "Concepts and results for 3D digital terrain-based wave propagation models: An overview," *IEEE J. Select. Areas Commun.*, vol. 11, no. 7, pp. 1002–1012, Sept. 1993.
- [24] W. C. Lee and D. J. Lee, "Pathloss predictions from microcell to macrocell," in *Proc. IEEE 51st Vehicle Technology Conf. (VTC-Spring)*, vol. 3, Tokyo, Japan: IEEE, May 2000, pp. 1988–1992.
- [25] J. H. Whitteker, "Physical optics and field-strength predictions for wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 515–522, Apr. 2002.
- [26] H. H. Xia, "A simplified analytical model for predicting path loss in urban and suburban environments," *IEEE Trans. Veh. Technol.*, vol. 46, no. 4, pp. 1040–1046, Nov. 1997.
- [27] S. Y. Tan and H. S. Tan, "Propagation model for microcellular communications applied to path loss measurements in Ottawa city streets," *IEEE Trans. Veh. Technol.*, vol. 44, no. 2, pp. 313–317, May 1995.
- [28] S. F. Hamim and M. F. Jamlos, "An overview of outdoor propagation prediction models," in *Proc. IEEE 2nd Int. Symp. Telecommunication Technologies (ISTT)*, Nov. 2014, pp. 385–388.
- [29] Infovista, "In wireless network planning, accuracy matters." Accessed on: Aug. 2019. [Online]. Available: www.infovista.com/products/Mentum-Planet-Live-RF-planning-and-optimization
- [30] Forsk, "Atoll overview." Accessed on: Aug. 2019. [Online]. Available: www.forsk.com/atoll-overview
- [31] N. Yuan, Z. Wei, X. Zhang, and D. Yang, "A new path loss predicting strategy for radio network planning," in *Proc. IEEE 68th Vehicle Technology Conf. (VTC-Fall)*, Sept. 2008, pp. 1–5.
- [32] E. Ostlin, H.-J. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2735–2747, July 2010.
- [33] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 2002.
- [34] G. R. MacCartney and T. S. Rappaport, "Rural macrocell path loss models for millimeter wave wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1663–1677, July 2017.

